Disease Markers 31 (2011) 247–257
DOI 10.3233/DMA-2011-0825
IOS Press

Exploratory investigation on functional significance of *ETS2* and *SIM2* genes in Down syndrome

Arpita Chatterjee, Samikshan Dutta*, Swagata Sinha and Kanchan Mukhopadhyay**
Manovikas Biomedical Research and Diagnostic Centre, MRIH, Kolkata, India

Abstract. Trisomy of the 21st chromosome leads to an over dosage of several regulatory genes in Down syndrome (DS). Though allelic and genotypic combinations formed between genes are interesting, till date, this particular area has never been explored in DS. In the present investigation four SNPs in two transcription factors, Single minded 2 (*SIM2*) and V-ets erythroblastosis virus E26 oncogene homolog2 (*ETS*2), located in the 21st chromosome were genotyped to understand their role in DS. Genomic DNA of eastern Indian probands with DS (N = 132), their parents (N = 209) and ethnically matched controls (N = 149) was subjected to PCR-based analyses of functionally important SNPs followed by statistical analyses. *ETS2* rs461155 showed high heterozygosity in DS. Significantly lower frequency of *SIM2* C-G haplotype (rs2073601-rs2073416) was noticed in individuals with DS (P value = 0.01669) and their fathers (P value = 0.01185). Significantly lower frequency of the A-C-C-G with higher frequency of A-C-A-G haplotypes was also noticed in subjects with DS (P value = 0.02089 and 0.00588 respectively). Data obtained indicate that the rs2073601 'A' allele, responsible for nonsynonymous substitution of leucine to methionine, may have some role in DS in this population.

Keywords: Down syndrome, ETS2, SIM2, transcriptional regulation

1. Introduction

Down syndrome (DS) (MIM# 190685) is the most common cause of intellectual disability throughout the world [1]. Along with intellectual disability compromised immunity, hormonal alteration, predisposition to malignancy etc. are of frequent occurrence in individuals with DS. Although trisomy of the 21st chromosome (HSA21) as a whole or triplication of genes in the Down syndrome critical region, leading to an over expression of these genes, may be one of the key factor for DS [2], it is still uncertain whether DS and its associated physiological abnormalities could be caused by particular gene loci in HSA21 [3].

Single minded 2 (*SIM2*) and V-ets erythroblastosis virus E26 oncogene homolog 2 (*ETS2*) are two transcription factors located in HSA21 and are over expressed in individuals with DS [1]. Downstream genes of these TFs are thought to play important roles in DS associated patho-physiology [1]. Further, higher differential expression ratio observed for the TFs in different tissues made them important candidates for disease association studies [4].

SIM2 belongs to basic-helix-loop-helix (bHLH), PER-ARNT-SIM (PAS) family of TFs [5–7]. It can heterodimerize with aryl hydrocarbon receptor nuclear translocator (ARNT) and migrate to the nucleus where bHLH domain binds to the consensus E-box {5′GT(G/A)CGTG 3′}of central nervous system (CNS) midline enhancer (CME) [8]. After binding to CME, the C terminal end of the bHLH-PAS domain activates the CNS midline gene transcription [9]. Development of nervous system in human depends on proper functioning of CNS midline cells and *CNS* midline gene expression requires midline precursor expression

*Current address: Dongguk University, School of Medicine, Dept. of Anatomy, Gyeongju, South Korea.
**Corresponding author: Kanchan Mukhopadhyay, Manovikas Biomedical Research and Diagnostic Centre, 482, Madudah, Plot I-24, Sec.-J, E.M. Bypass, Kolkata – 700107, India. Tel.: +91 033 4001 9179; Fax: +91 033 2442 8275; E-mail: kanchanmvk@yahoo.com; mvk_km@yahoo.co.in.

ISSN 0278-0240/11/$27.50

regulated by the single-minded gene product [5] making *SIM2* an important candidate for normal neuronal development. Expression of SIM2 mRNA has been detected in brain regions reported to be associated with DS pathology [10].

ETS2 is another TF located in HSA21 and identified to have important regulatory role in DS related abnormalities [1]. Over expression of *ETS2* was found to induce DS like craniofacial defects and skeletal anomalies in transgenic mice [11], to increased rate of neuronal apoptosis [12] and amyloid precursor protein (APP) gene transactivation. Over expression of *ETS2* was hypothesized to play vital role in the pathogenesis of early onset Alzheimer's disease and neuronal abnormalities in individuals with DS by regulating APP gene function [13].

The present study was aimed at investigating importance of four single nucleotide polymorphisms (SNPs), rs461155 (A/G) (coding, synonymous) and rs1051425 (C/T, 3′-UTR) SNPs in *ETS2* and rs2073601 (C/A) (coding, nonsynonymous) and rs2073416 (G/A) (coding, synonymous) in *SIM2*, in DS.

2. Materials and methods

2.1. Subjects

Peripheral blood was collected from ethnically matched healthy control individuals (N = 149) without having any clinical history of intellectual disability. Unrelated nuclear families (N = 132) with 78 male (59.09%) and 54 female (40.91%) DS probands, within an age range of 0.8–23 years, attending the out patient department of Manovikas Kendra, Kolkata, were recruited following the DSM-IV criteria [14]. Enlisted subjects consisted of 83 complete parent-offspring trios, 41 duos (35 without father, 6 without mother) and 8 DS probands. Study protocol was approved by the Institutional Human Ethical Committee.

After obtaining informed written consent, peripheral blood was collected from all the recruited individuals and leukocytes, cultured in presence of phytohemagglutinin, were used for karyotype analysis [15] for confirmation of diagnosis; all the subjects with DS recruited for this study showed trisomy 21 while parents and controls were devoid of any gross chromosomal abnormalities.

2.2. Selection of SNP

Risk conferred by individual SNPs was analyzed computationally by FastSNP (http://fastsnp.ibms.sinica.edu.tw) [16]. Functional significance of SNPs was analyzed by Pupasuite 2 (http://pupasuite.bioinfo.cipf.es) [17] which gives a score difference between two alleles for the SNP.

2.3. Genotyping

Leucocyte genomic DNA, isolated using standard protocol [18], was used for PCR-based amplification. Target sequences were subjected to sequencing in ABI prism 3130 Genetic Analyzer using Big Dye sequencing kit v3.1 followed by analysis using Sequencing Analysis software v 5.2. Details of the procedure are described in Table 1.

2.4. Statistical analyses

Allelic and genotypic distributions of the studied *SIM2* SNPs in eastern Indian control population (IND-C) were compared with that of four ancestral populations studied in the International HapMap project. Allele and genotype frequencies of controls were also compared with subjects with DS and their parents by simple r x c contingency table (http://www.physics.csbsju.edu/stats/contingency_NROW_NCOLUMN_form.html). Allelic odds ratios were calculated by Odds ratio calculator (http://www.hutchon.net/ConfidORnulhypo.htm).

Association between DS and *ETS2* rs461155 and rs1051425 were reported earlier [19]. In this investigation, genotype data was used for calculating LD, haplotypic association and interaction analysis with *SIM2* rs2073601 and rs2073416.

LD values were calculated by Haploview 4.1 using default settings. Haplotype frequency distributions of polymorphic SNPs were inspected by Unphased program (Version 2.404) [20]. Interaction among the genotypes of rs461155, rs1051425, rs2073601 and rs2073416 was analyzed by multifactor dimensionality reduction (MDR) version 2.0 beta 3 [21]; interaction graphs were used to calculate nature of dependencies or interactions using the MDR algorithm. The MDR interaction model describes percentage of entropy (information gain or IG) by each factor or 2-way interaction. Nodes and connections indicate the IG in case-control status removed by each variable (independent main effect) and by each pairwise combination of at-

Table 1
Analytical method used for studying *SIM2* and *ETS2* SNPs

SNP ID (Alleles)	Type of SNP (Amino acid change)	Primer sequence	Genotyping
ETS2 rs461155 (A/G)	Coding synonymous	F:5′GTTGTCTTTGCCAGGGACTC-3′ R:5′CGGTGAATGTGGTACTGTGG-3′	Analysis of PCR amplicons in ABI
ETS2 rs1051425 (T/C)	3′UTR SNP	F:5′CAAGGGCCGACTAAGAGAAG-3′ R:5′GCATGCAAAGAAGTGGAAAA-3′	3130 Genetic Analyzer using Big Dye kit v3.1 followed
SIM2 rs2073601 (C/A)	Coding NS (L483M)	F:5′ATCCATCTTCTTTGCCATGC-3′ R:5′GGTGGCTCTGGAGGATTTTT-3′	ow by sequence analysis using Sequencing software
SIM2 rs2073416 (G/A)	Coding synonymous		v5.2.

tributes (interaction effect) respectively; thus, independent main effect of each polymorphism can be quickly compared with the interactive effect.

Power of all the chi square tests was calculated by Piface program [22]. The SNPs are designated as 1 (rs461155), 2 (rs1051425), 3 (rs2073601) and 4 (rs2073416) for all the statistical analyses. Triplicate genotypes under homozygous conditions were considered as diploid homozygous genotypes in DS probands while the triplicate heterozygous genotypes were considered as the diploid heterozygous genotype for all the calculations to compare with respective reference diploid groups.

2.5. Prediction of nondisjunctioal status in DS probands

For obtaining information regarding parental and stage of origin of meiotic nondisjunction in probands with DS, pericentric STR markers are the ideal choice [23,24]. However, in absence of that information, in this study we have used a different approach for predicting parental and stage of origin of meiotic nondisjunction in individuals with DS. As was proposed by previous investigators [25], in case of homozygous proband of two heterozygous parents, the stage of nondisjunction was speculated as meiosis II. The reverse cases (heterozygous proband with one homozygous and one heterozygous parent) were considered as the case of nondisjunction at meiosis I. Information for all the four genetic loci was considered separately for predicting the stage of origin of nondisjunction. For families uninformative for a certain locus, nondisjunction status was ascertained from other informative loci.

3. Results

Among several SNPs present in the *SIM2*, FastSNP analysis showed that rs2073601 (C/A) [coding nonsynonymous (L483M)] and rs2073416 (A/G) [coding synonymous] may play deleterious role in the functioning of *SIM2* since both bear low to medium risk (risk level: 2–3). Analysis by Pupasuite 2 program revealed that both the SNPs can alter SR protein mediated splicing of SIM2 mRNA. rs461155 and rs1051425 in *ETS2* were also analyzed by Pupasuite 2 and changes in SRp40 mediated exonic splicing enhancer activity and transcriptional regulation activity respectively were noticed. These four SNPs were selected for analysis in the present study.

3.1. Allelic and genotypic frequency distribution pattern of SIM2 SNPs

Minor allele (A) frequency of rs2073601 (C/A) was significantly lower in the IND-C population (Fig. 1A) as compared to the Caucasians from Utah with ancestry from western and northern Europe (CEU) ($\chi^2 = 5.88$, P value $= 0.015$) while being significantly higher to that of the Yoruba from Ibadan, Nigeria (YRI), Han Chinese from Beijing, China (HCB) and Japanese from Tokyo, Japan (JPT) ($\chi^2 = 21.4$, 24.2, 27.3; P value < 0.0001 respectively). Minor allele (A) frequency of rs2073416 (G/A) in the IND-C population was found to be similar with HCB and JPT populations but different from the CEU ($\chi^2 = 6.45$, P value $= 0.011$) and YRI ($\chi^2 = 5.43$, 0.020) populations (Fig. 1A).

Comparison between different Indian subjects revealed that in probands with DS, frequency of the rs2073601 'A' allele was slightly higher ($\chi^2 = 3.61$, P value $= 0.057$) as compared to the IND-C population

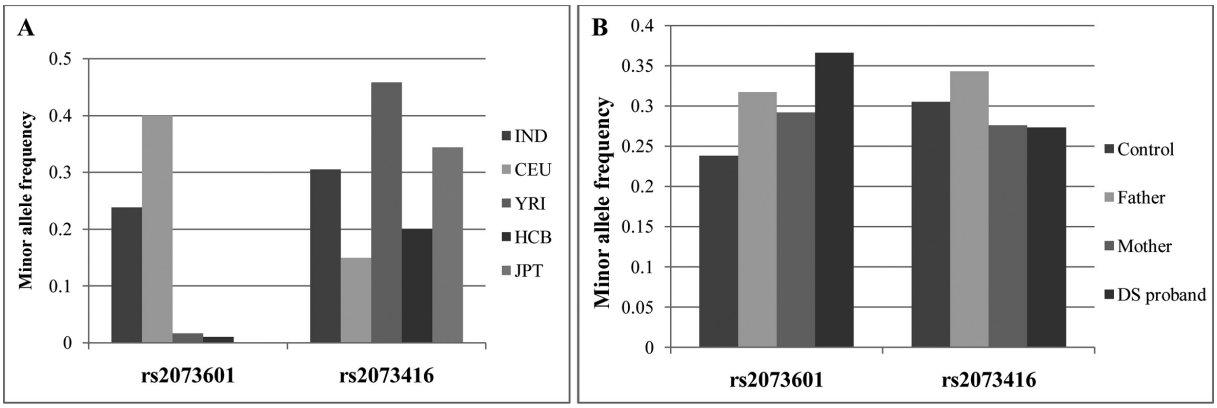

Fig. 1. [**A**] Frequency of rs2073601 (C/A) and rs2073416 (G/A) minor alleles in the IND-C and four ancestral populations [CEU, YRI, HCB and JPT] studied in the International HapMap project; [**B**] comparative analysis of minor allele frequencies in nuclear family members with DS probands and IND-C.

(Fig. 1B). Further, genotype distribution of rs2073601 showed significantly higher occurrence of CCA/CAA (0.515) and AAA (0.076) genotypes in DS with concomitant low (0.409) CCC genotype ($\chi^2 = 7.18$, P value = 0.028, power = 29.46%) as compared to the IND-C (CC = 0.557; CA = 0.409; AA = 0.034). rs2073416 did not show any significant difference in allelic and genotypic distribution (Fig. 1B).

3.2. Linkage disequilibrium and haplotype distribution pattern of ETS2 and SIM2 SNPs

The normalized LD (D′) and correlation coefficient (r^2) values for rs461155-rs1051425 and rs2073601-rs2073416 were comparatively higher in probands with DS ($\chi^2 = 42.5$, $P < 0.000$, power = 89.23%), their father ($\chi^2 = 12.0$, P value = 0.035, power = 47.34%) and mother ($\chi^2 = 17.5$, P value = 0.004, power = 52.76%) in comparison to the control group (Table 2). However, the r^2 values in all these groups were low and it can be concluded that the sites are in weak LD.

Haplotype analysis showed that frequency of the reference haplotype (A-C-C-G), containing ancestral allele of all the four SNPs, was significantly lower in probands with DS ($\chi^2 = 5.336$, P value = 0.02089) with concomitant higher frequency of the A-C-A-G haplotype ($\chi^2 = 7.587$, P value = 0.00588) (Fig. 2; best p value = 0.01017, global significance level = 0.0421, standard error = 0.002008). Independent analysis of rs2073601 and rs2073416 also showed lower frequency of C-G haplotype in probands with DS (P = 0.01669) and their fathers (P = 0.01185).

3.3. Interaction of ETS2 and SIM2 genotypes

Either the wild type homozygous genotypes and/or heterozygous genotypes of these three SNPs were present in higher frequencies in individuals with DS as compared to controls. Negligible co-occurrence of homozygous risk genotypes of these SNPs was noticed in the disease group.

Gene-gene interaction analysis revealed stronger independent effect of rs2073416 and rs2073601 (positive IG values in the nodes) in parents of subjects with DS while in the probands major effect of rs461155 was noticed (Fig. 3). Negative entropy observed for pairwise interactions indicated redundancy between these SNPs in all the groups (Fig. 3). Verification of these findings by cross-validation statistics of MDR (Table 3) revealed significant correlation in individuals with DS (rs461155-rs2073601 Testing Balance Accuracy, TBA = 0.5987, Cross Validation Consistency, CVC = 10, P = 0.0375; rs461155–rs2073601–rs2073416 TBA = 0.6062, CVC = 9, P = 0.0195) while in the parent groups, level of redundancy was higher with nominal CVC and P values (Table 3).

3.4. Status of nondisjunction in DS probands

Analysis of nondisjunctional status in probands with DS (Table 4) revealed that total number of paternal nondisjunction was 24, 12, 11 and 12 (total 59) while total number of maternal nondisjunction was 27, 21, 27 and 33 (total 108) for rs461155, rs1051425, rs2073601 and rs2073416 respectively. Out of these, nondisjunction at meiosis I possibly occurred 65 times whereas nondisjunction at meiosis II took place only 17 times

Table 2

Estimate of linkage disequilibrium (LD) parameters D′ (upper) and r^2 (lower) for the studied SNPs in ETS2 and SIM2 genes

SNPs	Control				Father				Mother				DS proband			
	rs461155	rs1051425	rs2073601	rs2073416	rs461155	rs1051425	rs2073601	rs2073416	rs461155	rs1051425	rs2073601	rs2073416	rs461155	rs1051425	rs2073601	rs2073416
rs461155	–	0.697	0.037	0.245	–	0.788	0.006	0.081	–	0.62	0.045	0.192	–	0.808	0.374	0.091
rs1051425	0.164	–	0.072	0.188	0.182	–	0.047	0.201	0.102	–	0.098	0.001	0.211	–	0.081	0.045
rs2073601	0.001	0.003	–	0.895	0	0	–	1	0.001	0.004	–	0.749	0.047	0.001	–	0.937
rs2073416	0.016	0.003	0.11	–	0.005	0.004	0.239	–	0.02	0	0.087	–	0.006	0.001	0.198	–

Table 3
Summary of the Multi Dimensionality Reduction cross-validation statistics

Groups	Best combination of SNP* in each dimension	TBA[a]	CVC[b]	P.E.[c] (1-TBA)	P Value of TBA
Father of DS proband	3	0.5321	9	0.4679	0.6020
	1–3	–	–	–	–
	3–4	0.5732	8	0.4268	0.2500
	1-3-4	0.6010	9	0.399	0.0695
	1-2-3-4	0.5679	10	0.4321	0.2865
Mother of DS proband	3	0.5200	7	0.4800	0.6790
	1–3	0.5472	6	0.4528	0.4240
	3–4	0.5472	6	0.4528	0.4240
	1-3-4	0.5817	9	0.4183	0.1075
	1-2-3-4	0.5954	10	0.4046	0.0525
DS Proband	1	0.5765	8	0.4235	0.1350
	1–3	**0.5987**	**10**	**0.4013**	**0.0375**
	1-3-4	**0.6062**	**9**	**0.3938**	**0.0195**
	1-2-3-4	0.5531	10	0.4469	0.3545

*1 = rs461155, 2 = rs1051425, 3 = rs2073601, 4 = rs2073416; [a]Testing balance accuracy; [b]Cross validation consistency; [c]Prediction error.

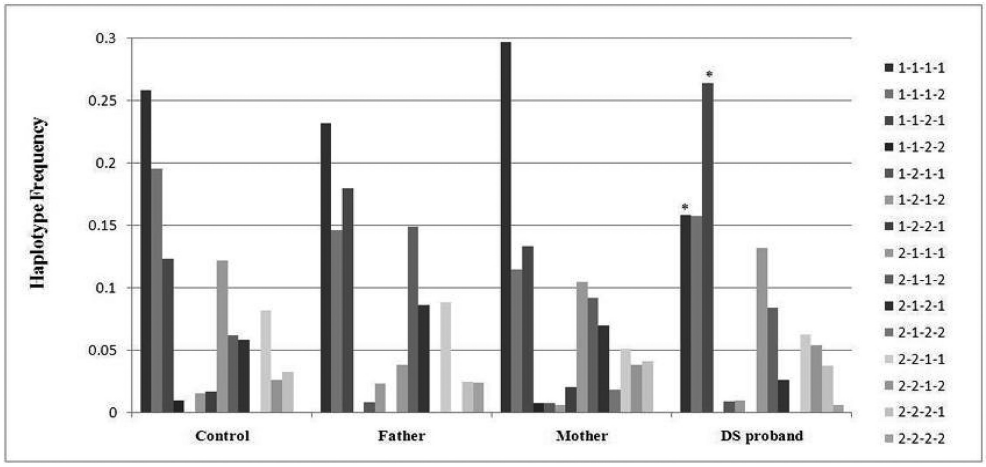

Fig. 2. Frequency of haplotypes formed between four SNPs in probands with DS and their family members. * haplotypes (A-C-C-G and A-C-A-G) showing significant difference in DS probands as compared to controls (P value = 0.02089 and 0.00588 respectively).

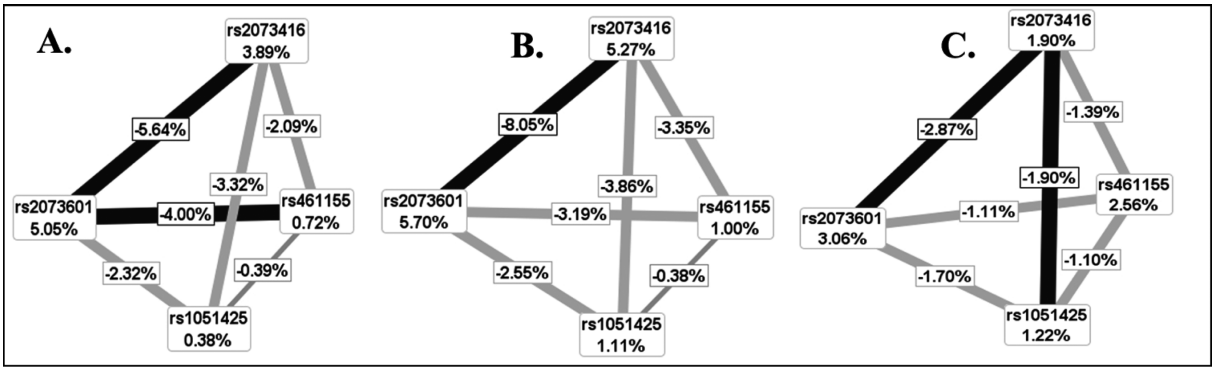

Fig. 3. Independent effect of SNPs (positive IG values in the nodes) and their interactions (lines between two SNPs) obtained by MDR 2.0 beta 3 in fathers (A), mothers (B) and probands with DS (C). Redundancy is depicted as a line between factors accompanied by a negative percent of entropy.

Table 4
Predicted stage of origin of nondisjunction

SNP ID	Paternal origin of nondisjunction			Maternal origin of nondisjunction		
	Meiosis I	Meiosis II	Meiosis I/II	Meiosis I	Meiosis II	Meiosis I/II
rs461155	11	1	12	11	4	12
rs1051425	4	0	8	6	2	13
rs2073601	2	3	6	12	3	12
rs2073416	7	1	4	12	3	18

Table 5
Genotypic status of rs461155 in DS probands and their parents

NDJP* genotype	CDJP** genotype	Proband genotype	Status of Marker§	Predicted stage of NDJ	Trios/ Duos /Singleton
AA ♂/♀	GG ♂/♀	AAG	N	I/II	6
AA ♂/♀	AG ♂/♀	AAA	R	I	6
AG ♂/♀	AA ♂/♀	AAG	N	I	15
AG ♂/♀	AA ♂/♀	AGG	N	II	2
AG ♂/♀	AG ♂/♀	GGG	R	II	1
AG ♂/♀	AG ♂/♀	AAA	R	II	4
AG ♂/♀	GG ♂/♀	AGG	N	I	7
GG ♂/♀	AG ♂/♀	GGG	N	I	3
AG ♀	GG ♂	AAG	N	II	1
AG ♂	AG ♀	AAG	N	I/II	15
AG ♂	AG ♀	AGG	N	I/II	14
missing ♀	AA ♂	AGG	N	I/II	1
missing ♀	GG ♀	AAG	N	I/II	1
missing ♀	AG ♂	AGG	N	I/II	1
missing ♀	AA ♂	AAA	R	I/II	1
missing ♀	AG ♂	AAA	R	I/II	1
missing	AG ♂	GGG	R	I/II	2
missing ♂	AG ♀	AAA	R	I/II	3
missing ♂	AA ♀	AAA	R	I/II	5
missing ♂	AG ♀	AGG	N	I/II	2
missing ♂	AA ♀	AAG	N	I/II	4
missing ♂	GG ♀	AGG	N	I/II	4
missing ♂	AG ♀	AAG	N	I/II	16
missing	missing	AAA	R/N	I/II	4
missing	missing	AAG	N	I/II	1
missing	missing	GGG	R/N	I/II	1
missing	missing	AGG	N	I/II	2

*NDJP: Nondisjoining parent.
**CDJP: Correctly disjoining parent.
§R: reduced to homozygocity, N: Not reduced to homozygocity.

($\chi^2 = 53.2$, $P < 0.0001$). From the above data it can be assumed that overall rate of maternal nondisjunction was higher than paternal for the studied sites with a significantly higher occurrence of nondisjunction at meiosis I ($\chi^2 = 8.89$, $P = 0.003$).

Since rs461155 and rs2073601 showed strong interaction and significant difference in allelic frequencies in subjects with DS as compared to control, allelic transmission pattern of these two sites was analyzed in detail. Genotypic analysis reveled that in total, 8 and 24 families were uninformative for rs461155 and rs2073601 respectively. An overdosage (61.16%) of

Table 6
Genotypic status of rs2073601 in DS probands and their parents

NDJP* genotype	CDJP** genotype	Proband genotype	Status of Marker§	Predicted stage of NDJ	Trios/ Duos /Singleton
CC ♂/♀	AA ♂/♀	CCA	N	I/II	2
CC ♂/♀	CA ♂/♀	CCC	R	I	10
CA ♂/♀	CC ♂/♀	CCA	N	I	9
CA ♂/♀	CC ♂/♀	CAA	N	II	8
CA ♂/♀	CA ♂/♀	AAA	R	II	3
CA ♂/♀	CA ♂/♀	CCC	R	II	1
CA ♂/♀	AA ♂/♀	CAA	N	I	4
CC ♂/♀	AA ♂/♀	CAA	N	I/II	7
AA ♀	CA ♂	AAA	N	I	3
CA ♂	CA ♀	CCA	N	I/II	2
CA ♂	CA ♀	CAA	N	I/II	10
missing ♀	AA ♂	CAA	N	I/II	1
missing ♀	CA ♂	CCA	N	I/II	1
missing ♀	CA ♂	CAA	N	I/II	1
missing ♀	CC ♂	CCC	R	I/II	2
missing ♀	CA ♂	CCC	R	I/II	1
missing ♂	CA ♀	CCC	R/N	I/II	2
missing ♂	CC ♀	CCC	R	I/II	11
missing ♂	CA ♀	CAA	N	I/II	7
missing ♂	CC ♀	CCA	N	I/II	4
missing ♂	AA ♀	CAA	N	I/II	2
missing ♂	AA ♀	AAA	R	I/II	2
missing ♂	CA ♀	CCA	N	I/II	3
missing ♂	CA ♀	AAA	R	I/II	1
missing ♂	CC ♀	CAA	N	I/II	3
missing	missing	CCC	R/N	I/II	2
missing	missing	CCA	N	I/II	1
missing	missing	AAA	R/N	I/II	1
missing	missing	CAA	N	I/II	4

*NDJP: Nondisjoining parent.
**CDJP: Correctly disjoining parent.
§R: reduced to homozygocity, N: Not reduced to homozygocity.

the "A" allele of rs461155 was noticed in probands with DS (AAG-59, AAA-24) in comparison to the dosage of the "G" allele (GGG-7; AGG-33). Among 124 families with 189 parents (74 with both parents and 41 with single parent) the "G" allele was predicted to disjoin correctly in 86 instances while the "A" allele disjoined correctly 96 times (Table 5). We have also noticed that reduction to homozygosity was low for this site; out of 122 probands recruited with parents, marker status was reduced to homozygosity in only 23 cases.

Analysis of 108 informative cases with DS for rs2073601 (Table 6) revealed that the CAA genotype was predominant among probands (43.5%). Frequency of probands with AAA genotype (9.26%) was significantly low as compared to the CCC genotype (26.85%). On the other hand, more than 73% cases contained the "A" allele (10 with AAA, 47 with AA and 22 with A while 29 was without A). Among 122 probands with informative families reduction to homozygosity was noticed in 31 probands.

4. Discussion

In the present investigation, for the first time haplotypic association of *SIM2* and *ETS2*, two TFs located

in the DSCR, have been analyzed in probands with DS. Among the four SNPs studied in *ETS2* and *SIM2*, the nonsynonymous rs2073601 exhibited significantly altered frequency distribution pattern in comparison to all the four populations enlisted in HapMap. Differences in allelic frequencies between populations have been predicted to have good correlation with differences in gene expression [26]. We have observed that allelic change of rs2073601 may lead to a nonsynonymous substitution; while the "C" allele codes for leucine, a basic amino acid, the 'A' substitution leads to encode for methionine, a nonpolar amino acid. This may induce a loss in SR protein mediated activities (SC35 activity score = C-2.93, A-1.66, lose −1.27; SF2 activity score = C-3.74, A-1.22, loss-2.52). Our investigation has also revealed a higher frequency of the derived "A" allele and "AA" genotype of rs2073601 in eastern Indian probands with DS. It has been proposed that SR proteins can affect both constitutive as well as regulated splicing [27] and SNPs in the binding site may alter function of the protein [28]. Therefore, from the data obtained in the present study it may be hypothesized that eastern Indian probands with DS, with a higher frequency of the rs2073601 "A" allele, has higher chances of harboring SIM2 splice variants which may lead to an alteration in protein function.

rs461155 in *ETS2* also showed a nominal difference in allelic frequencies between probands with DS and control (χ^2 = 0.213E-01, P = 0.884). The "A" allele, playing a protective role, was found in excess in probands with DS (0.629) and from the data obtained it seems that this SNP may not have an independent role in the etiology of DS.

The heterozygous genotype of rs2073601 (i.e. 'CCA' or 'CAA') was frequently found to be present in DS subjects with the wild type or heterozygous genotype of other two SNPs (rs461155 and rs2073416). Since genotypic frequency distribution of rs2073601 revealed that the frequency of 'CAA' genotype (0.435) was much higher than that of 'CCA' genotype (0.204), it can be predicted that 'CAA' genotype could be interacting with other genotypes. However, this could be actually redundant in nature since gene-gene interaction analysis by MDR revealed redundant interaction of these sites in subjects with DS as well as their parents.

It has already been reported that in majority of cases with DS, the error occurs during maternal meiotic cycle [29] and more precisely during maternal meiosis I [24,30,31]. Similar observations were reported for the GluK1 kinate receptor in the eastern Indian population, another gene located at the DSCR [25]. In the present investigation also, analysis of genotypic data obtained from four SNPs revealed higher occurrence of maternal meiosis I nondisjunction in cases with DS (χ^2 = 8.89, P value = 0.003).

Major limitation of the present study is use of markers which cannot predict the actual stage of nondisjunction. However, use of four different dimorphic variations helped in deducing the stage of origin by comparative analyses. This preliminary investigation has shown a possibility of significant involvement for the rs2073601 with DS and further analyses involving large number of samples along with pericentric markers, may help in understanding the status of nondisjunction and actual risk involved with the proposed marker.

Characteristic features of DS develop mostly during the early growth period [30,31] and trisomy of the 21st chromosome generated by parental meiotic nondisjunction was hypothesized as the major cause [32]. An error in cell division thus may lead to a disturbance in dosage of certain genes controlling regulatory function and as a result, depending on the extent of the trisomic region, several important cellular mechanisms could be disrupted [23]. It can be speculated from the present investigation that the observed higher frequency of rs2073416 "A" allele may help in generation of *SIM2* splice variants and can induce malfunctioning of several downstream genes like hedgehog involved in cell linage specific development of CNS [33], or expression of neuroendocrine genes like thyroxin releasing hormone, somatostatin etc. [34]. These genes regulate early brain development and any adulteration in their function can damage normal brain development or intellectual functioning [33,34]. From the current preliminary investigation, an in depth study is warranted on the role of *SIM2* and its downstream genes in the patho-physiology of DS.

5. Conclusion

Important observations made during the current study are: 1) rs461155 and rs2073601, in *ETS2* and *SIM2* genes respectively, showed significant redundant interaction in probands with DS as compared to ethnically matched controls; the parent population failed to show such significance level, 2) both these sites showed significant differences in allelic frequencies as compared to the control population, 3) analysis of status of nondisjunction revealed higher maternal mieotic I nondisjunction which resembles that reported by earlier investigators [23,31]; 4) the rs2073601 "A" allele,

that showed significant difference in frequency, confer changes in nucleotide sequence leading to alteration in mRNA splicing and may contribute to generation of splice variants, 5) higher frequency of rs2073601 CAA genotype along with high occurrence of A-C-A-G haplotype in DS probands indicate that rs2073601 may serve as an important candidate for disease association studies particularly in the eastern Indian population.

Acknowledgements

Authors are thankful to all the individuals for their participation in the study. Indian Council of Medical Research is also acknowledged for providing senior research fellowship to AC (#45/1/2010-Hum/BMS).

References

[1] M. Rachidi and C. Lopes, Mental retardation in Down syndrome: From gene dosage imbalance to molecular and cellular mechanisms, *Neuroscience research* **59** (2007), 349–369.

[2] I. Kola and P.J. Hertzog, Animal models in the study of the biological function of genes of human chromosome 21 and their role in the pathophysiology of Down's syndrome, *Human Molecular Genetics* **6** (1997), 1713–1727.

[3] J.O. Korbel, T. Tirosh-Wagner, A.E. Urban, X.-N. Chen, M. Kawoski, L. Dai et al., The genotypic architecture of Down syndroem phenotypes revealed by high-resolution analysis of human segmental trisomies, *Proc Natl Acad Sci USA* **106** (2009), 12031–12036.

[4] R. Chen, A.A. Morgan, J. Dudley, T. Deshpande, L. Li, K. Kodama et al., FitSNPs: Highly differentially expressed genes are more likely to have variants associated with disease, *Genome Biology* **9** (2008), R170.

[5] K.A. Wharton, Jr., R.G. Franks, Y. Kasai and S.T. Crews, Control of CNS midline transcription by asymmetric E-box-like elements: similarity to xenobiotic responsive regulation, *Development* **120** (1994), 3563–3569.

[6] J.R. Nambu, J.O. Lewis, K.A. Jr. Wharton and S.T. Crews, The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development, *Cell* **67** (1991), 1157–1167.

[7] M.P. Ward, J.T. Mosher and S.T. Crews, Regulation of bHLH–PAS protein subcellular localization during Drosophila embryogenesis, *Development* **125** (1998), 1599–1608.

[8] J. Michaud and C.M. Fan, Single-minded–two genes, three chromosomes, *Genome Research* **7** (1997), 569–571.

[9] H.I. Swanson, W.K. Chan and C.A. Bradfield, DNA Binding Specificities and Pairing Rules of the Ah Receptor, ARNT, and SIM Proteins, *Journal of Biological Chemistry* **270** (1995), 26292–26302.

[10] A. Yamaki, S. Noda, J. Kudoh, N. Shindoh, H. Maeda, S. Minoshima, K. Kawasaki, Y. Shimizu and N. Shimizu, The mammalian single-minded (SIM) gene: mouse cDNA structure and diencephalic expression indicate a candidate gene for Down's syndrome, *Genomics* **35** (1996), 136–143.

[11] S.H. Sumarsono, T.J. Wilson, M.J. Tymms, D.J. Venter, C.M. Corrick, R. Kola, M.S. Lahoud, T.S. Papas, A. Seth and I. Kola, Down's syndrome–like skeletal abnormalities in Ets2 transgenic mice, *Nature* (London) **379** (1996), 534–538.

[12] E.J. Wolvetang, O.M. Bradfield, T. Hatzistavrou, P.J. Crack, J. Busciglio, I. Kola and P.J. Hertzog, Overexpression of the chromosome 21 transcription factor Ets2 induces neuronal apoptosis, *Neurobiology of Disease* **14** (2003), 349–356.

[13] E.W. Wolvetang, O.M. Bradfield, M. Tymms, S. Zavarsek, T. Hatzistavrou, I. Kola and P.J. Hertzog, The chromosome 21 transcription factor ETS2 transactivates the beta-APP promoter: implications for Down syndrome, *Biochim Biophys Acta* **1628** (2003), 105–110.

[14] American Psychiatric Association (1994) Diagnostics and Statistical Manual of Mental Disorders, 4th ed. (DSM-IV). Committee on Nomenclature and Statistics, Washinton, DC: American Psychiatric Association.

[15] N.C. Sun, E.H.Y. Chu and C.C. Chang, Staining method for the banding patterns of human mitotic chromosomes, *Caryologia* **27** (1974), 315.

[16] H.Y. Yuan, J.J. Chiou, W.H. Tseng, C.H. Liu, C.K. Liu, Y.J. Lin et al., FASTSNP: an always uptodate and extendable service for SNP function analysis and prioritization, *Nucleic Acids Research* **34** (2006), W635–W641.

[17] L. Conde, J.M. Vaquerizas, H. Dopazo, L. Arbiza, J. Reumers, F. Rousseau et al., PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes, *Nucleic Acids Research* **34** (2006), W621–W625.

[18] S.A. Miller, D.D. Dykes and H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells, *Nucleic Acids Research* **16** (1988), 1215.

[19] A. Chatterjee, S. Dutta, S. Mukherjee, N. Mukherjee, S. Chandra, A. Mukherjee, S. Sinha, C.K. Panda, K. Chaudhuri and K. Mukhopadhyay, Differential allelic distribution of V-ets erythroblastosis virus E26 oncogene homolog2 (ETS2) functional polymorphisms in different group of patients, *Gene Expression* **15** (2011), 61–73.

[20] F. Dudbridge, Pedigree disequilibrium tests for multilocus haplotypes, *Genetic Epidemiology* **25**(2) (2003), 115–121.

[21] H. Mei, D. Ma, A. Ashley-Koch and E.R. Martin, Extension of multifactor dimensionality reduction for identifying multilocus effects in the GAW14 simulated data, *BMC Genetics* **6** (2005), S145–S149.

[22] R.V. Lenth, Statistical power calculations, *Journal of Animal Science* **85** (2007), E24–E29.

[23] Z. Xu, K.F. Kerstann, S.L. Sherman, A. Chakravarti and E. Feingold, A Trisomic Transmission Disequilibrium Test, *Genetic Epidemiology* **26** (2004), 125–131.

[24] S. Ghosh, E. Feingold and S.K. Dey, Etiology of Down syndrome, Evidence for consistent association among altered meiotic recombination, nondisjunction, and maternal age across populations, *American Journal of Medical Genetics Part A* **149A** (2009), 1415–1420.

[25] D. Ghosh, S. Sinha, A. Chatterjee and K. Nandagopal, A study of GluK1 kinate receptor polymorphisms in Down syndrome reveals allelic non-disjunction at 1173 (C/T), *Disease Markers* **27** (2009), 45–54.

[26] R.S. Spielman, L.A. Bastone, J.T. Burdick, M. Morley, W.J. Ewens and V.G. Cheung, Common genetic variants account for differences in gene expression among ethnic groups, *Nature Genetics* **39** (2007), 226–231.

[27] X.D. Fu, A. Mayeda, T. Maniatis and A. Krainer, General splicing factors SF2 and SC35 have equivalent activities in

vitro, and both affect alternative splicing alternative 5'and 3' splice site selection, *PNAS USA* **89** (1992), 11224–11228.

[28] M. Gabut, M Mine, C. Marsac, M. Brivet, J. Tazi and J. Soret, Pyruvate Dehydrogenase mRNA in a Case of Mental Retardation with Lactic Acidosis, *Molecular and Cellular Biology* **25** (2005), 3286–3294.

[29] S.L. Sherman, E.G. Allen, L.H. Bean and S.B. Freeman, Epidemiology of Down syndrome, *Mental Retardation and Developmental Disabilities Research Review* **13**(3) (2007), 221–227.

[30] S.E. Antonakaris, Parental origin of the extra chromosome in trisomy 21 as indicated by analysis of DNA polymorphisms, Down Syndrome Collaborative Group, *New England Journal of Medicine* **324** (1991), 872–876.

[31] S.E. Antonarakis, M.B. Petersen, M.G. McInnis, P.A. Adelsberger, A.A. Schinzel, F. Binkert et al., The meiotic stage of nondisjunction in trisomy 21: determination by using DNA polymorphisms, *American Journal of Human Genetics* **50**(3) (1992), 544–550.

[32] E. Niebbuhr: Down syndrome the possibility of a pathogenic segment of chromosome no. 21, *Humangenetik* **21** (1974), 99–101.

[33] M. Rachidi, C. Lopes, G. Charron, A.L. Delezoide, E. Paly, B. Bloch and J.M. Delabar, Spatial and temporal localization during embryonic and fetal human development of the transcription factor SIM2 in brain regions altered in Down syndrome, *International Journal of Developmental Neuroscience* **23** (2005), 475–484.

[34] E. Goshu, H. Jin, J. Lovejoy, J. Franc, O. Marion, J.L. Michaud and C.M. Fan, Sim2 Contributes to Neuroendocrine Hormone Gene Expression in the Anterior Hypothalamus, *Molecular Endocrinology* **18**(5) (2004), 1251–1262.